

基于矩阵填充问题的五轮零知识身份认证方案

王后珍^{1,2}, 蔡鑫伟¹, 郭岩¹, 张焕国¹

(1. 武汉大学国家网络安全学院, 湖北 武汉 430072; 2. 密码科学技术国家重点实验室, 北京 100878)

摘要: 针对现存绝大多数身份认证协议容易遭受量子计算攻击及实现效率低的缺陷, 基于矩阵填充 (MC) 问题构造了一种安全高效的五轮零知识身份认证方案。由于 MC 问题是 NP 完全的, 所提方案具有很好的抗量子计算攻击潜力。相较于目前已有类似方案, 所提方案通过增加单轮交互将欺骗概率由 $2/3$ 降至 $1/2$, 同时兼具容易实现、密钥尺寸小等优点。此外, 采用 Fiat-Shamir 密码转换技术还可将所提五轮零知识认证协议转换为高效的具有抗量子计算攻击潜力的数字签名方案。

关键词: 抗量子计算密码; 身份认证; 零知识证明; 矩阵填充问题

中图分类号: TP3-0

文献标识码: A

DOI: 10.11959/j.issn.1000-436x.2021212

5-pass zero-knowledge identity authentication scheme based on matrix completion problem

WANG Houzhen^{1,2}, CAI Xinwei¹, GUO Yan¹, ZHANG Huanguo¹

1. School of Cyber Science and Engineering, Wuhan University, Wuhan 430072, China

2. State Key Laboratory of Cryptology, Beijing 100878, China

Abstract: To solve the problem that most identity authentication schemes are vulnerable to quantum-computing attacks and low efficiency, a new 5-pass zero-knowledge identity authentication scheme was designed based on the matrix completion problem (MCP). Since the MCP is NP-complete, the proposed scheme has the potential to avoid quantum-computing attacks. Compared with the existing similar protocols, the proposed scheme reduced the fraud probability from $2/3$ to $1/2$ by adding a single round of interaction, and had the advantages of easy implementation and small key size. Moreover, based on the proposed zero-knowledge authentication scheme and Fiat-Shamir standard transformation method, a secure and efficient digital signature algorithm against quantum-computing can be obtained.

Keywords: post-quantum cryptography, identity authentication, zero-knowledge proof, matrix completion problem

1 引言

Goldwasser、Micali 等^[1]给出了零知识身份认证的定义,其含义是 P 试图使 V 相信其掌握某个知识,或证明论断的正确性,但在该过程中 V 或第三方无法获得任何与知识相关的内容。整个交互过程的关键即如何做到使 V 无法获取与知识本身相关的任何

信息。自从身份认证协议概念提出后,先后涌现了大批零知识身份认证协议,其中最具代表性协议包括 FS 协议^[2]、GQ 协议^[3]以及 Schnorr 协议^[4]等。它们的安全性主要基于数论困难问题,如大整数因子分解问题 (IFP, integer factorization problem) 或有限域上的离散对数问题 (DLP, discrete logarithm problem)。

收稿日期: 2021-07-31; 修回日期: 2021-10-31

基金项目: “十三五”国家密码发展基金资助项目 (No.MMJJ201701304); 国家自然科学基金资助项目 (No.61332019); 国家重点研发计划基金资助项目 (No.2018YFC1604000)

Foundation Items: The National Cryptography Development Fund of China (No.MMJJ201701304), The National Natural Science Foundation of China (No.61332019), The National Key Research and Development Program of China (No. 2018YFC1604000)

Shor^[5]在 1994 年设计了一种能够以多项式时间复杂度求解 IFP 和 DLP 的量子专用算法,也就是说,一旦造出可实用的量子计算机,基于上述数学困难问题设计的密码算法容易遭受攻击。现今密码学领域越来越重视抗量子计算密码的研究与探索,而且一些国家或组织正在积极开展抗量子计算密码算法的标准化工作^[6]。

目前,已有的抗量子身份认证密码方案主要如下。1993 年美密会上, Stern^[7]基于 Syndrome Decoding 问题构造的认证密码方案。后来,文献[8-9]在 Stern 方案的基础上做了进一步的优化,减小了 Stern 方案的公钥尺寸,减少了交互过程中的数据传递量,并提升了协议的安全性。上述方案均是在有限域 \mathbb{F}_2 上设计的,而文献[10]则将上述方案拓展到有限域 \mathbb{F}_q 上 (q 为素数)。2001 年亚密会上, Courtois^[11]基于矩阵最小秩问题提出零知识认证协议; 2011 年美密会上, Sakumoto 和 Shirai 等^[12]基于 MQ 问题成功设计出安全实用的零知识认证协议。其他类似的方案有 PKP 协议^[13]、Chen 协议^[14]、CLE 协议^[15]、PPP 协议^[16]、SC 协议^[17]等。上述身份认证协议单轮的欺骗概率主要介于 $1/2 \sim 3/4$, 因此,工程应用中需要通过多轮迭代才能满足实际期望的安全性。2019 年, Yang 等^[18]基于格上困难问题构造了一种欺骗概率可达到 $1/\text{poly}$ 的认证协议,不需要多轮迭代。2021 年,文献[19]借鉴 Courtois 方案的构造方法、基于矩阵填充 (MC, matrix completion) 问题构造了一种新型三轮零知识身份认证密码方案,相较于 Courtois 方案,该方案在欺骗概率不变的情况下,减小了密钥尺寸,并提升了协议的运行效率。本文的主要工作如下。

1) 在文献[19]中认证协议的基础上,进一步将恶意证明者欺骗成功的概率由 $2/3$ 降至 $1/2$,提出了基于低秩 MC 问题的五轮零知识身份认证方案。

2) 针对新设计的五轮零知识身份认证协议,本文给出了其完备性、合理性以及零知识性的详细证明,证明了方案的安全性。

3) 最后给出了本文方案同其他现有方案的参数比较,指出本文方案所具备的优势。

2 预备知识

这里特别说明,本文的零知识身份认证密码协议均在有限域 \mathbb{F}_q 上进行。其中, $q = p^h$, h 为正整数, p 为素数。

2.1 零知识身份认证协议

在日常生活中经常会遇到这样的问题,一方 P 希望向另一方 V 证明他具备某种身份或知道某个知识,通常习惯称 P 为证明者,称 V 为验证者,这是基本的身份认证的问题。但为了进一步确保方案的安全性,人们希望该过程中 V 或第三方无法获得任何与知识相关的内容,也就是平时所说的零知识身份认证方案。

当方案分别满足以下 3 个性质时,通常称该零知识身份认证方案是安全的。

完备性 (completeness)。如果陈述为真,那么诚实的证明者在不违背交互协议的前提下,总能使诚实的验证者相信其确实拥有知识。

合理性 (soundness)。如果陈述为假,那么恶意证明者几乎无法通过身份验证,诚实的验证者以绝对优势的返回拒绝,即恶意证明者成功通过验证的概率是可忽略的。下面给出正式定义。

定义 1 对任意概率多项式时间 (PPT, probabilistic polynomial time) 算法、验证者 P 以及输入 x , 存在提取器 δ , 如果有

$$\Pr[P, V(x)] > \delta_s + \varepsilon$$

其中, δ_s 为欺骗概率, ε 是不可忽略的, 则提取器 δ 可在多项式时间内获得相应有效信息, 且通过验证。

对于合理性这一性质有进一步的延伸, 也就是特殊合理性, 此处本文也对特殊合理性做简单说明。特殊合理性是指对于给定的公钥 pk , 当 $c_1 \neq c_2$ 时, 输出相同初始信息 I 的 2 个可接收记录 (I, c_1, r_1) 和 (I, c_2, r_2) 是困难的。也就是说, 对于任意的多项式时间敌手, 其做出的响应只能通过挑战值中的一个。下面同样给出特殊合理性的定义。

定义 2 如果能够证明身份认证协议具备特殊合理性, 那么对算法 A 而言, 只要是 PPT 算法, 如下概率便可忽略不计。

$$\Pr[(I, c_1, r_1, c_2, r_2) \leftarrow A(pk):$$

$$c_1 \neq c_2, \text{Rsp}(c_1, r_1) = \text{Rsp}(c_2, r_2) = \text{accept}] \leq \varepsilon$$

其中, (pk, sk) 由密钥算法生成, Rsp 表示对应记录验证者给出的判断。

零知识性 (zero-knowledge)。当证明者真正拥有知识 (私钥) 信息时, 诚实验证者或攻击者只要遵守协议运行规则, 无论采用什么方法, 如何对交

互过程中的传输数据进行推导, 均不能得到与知识本身相关的任何有效信息。下面给出对零知识性更规范定义。

定义 3 假如存在有效的模拟器 U , 对于验证者 V 和证明者 P 而言, 以下二者是计算上不可区分的:

1) $\{P(\text{sk}), V(\text{pk})\}$ 表示 V 和 P 在忠实执行交互后 V 所得到的知识信息, 其中, sk 为私钥, pk 为公钥;

2) $\{U(\text{pk})\}$ 表示输入为公钥 pk 、算法 U 的输出。

则称该身份识别协议是诚实验证者零知识的。

另外需要指出, 如果上述二者的分布是相同的, 那么称该协议满足完美零知识性。如无特殊说明, 本文后续的协议均指计算不可区分。为方便描述, 上述完备性和合理性定义中的陈述均指证明者拥有知识这一断言。

2.2 相关 NP 困难问题

下面分别简单描述矩阵最小秩 (MR, min rank) 问题和矩阵填充问题。

文献[20]较详细地描述了矩阵秩相关问题求解算法, 矩阵最小秩问题就是其中之一, 其求解思路是构造多变量非线性方程组, 然后解方程组。另外, 文献[21]也对 MR 问题进行了详细分析。下面首先给出 MR 问题的正式定义。

定义 4 (最小秩问题) 设 M_0, M_1, \dots, M_m 为有限域 \mathbb{F}_q 上的 $m \times n$ 矩阵, 寻找一个向量 $\alpha = (\alpha_1, \dots, \alpha_m) \in \mathbb{F}_q^m$ 使矩阵 $\sum \alpha_i M_i - M_0$ 的秩满足

$$\text{rank}(\sum \alpha_i M_i - M_0) \leq r, 1 \leq r < n$$

与矩阵最小秩问题相似的另一个数学困难问题——矩阵填充是本文重点关注的。文献[22-24]给出了矩阵填充问题是 NP 困难问题的证明, 特别地, 低秩矩阵填充问题的可表述如下

定义 5 (低秩矩阵填充问题) 随机给定一个 $\eta \times n$ 维矩阵 $A = (a_{ij}) \in \mathbb{F}_q$ 且 A 的秩 $\text{rank}(A) = r$, 假如随机去掉矩阵中的 m 个元素。现对这 m 个空缺位置进行赋值, 满足赋值后矩阵 A' 的秩仍然等于 r 。

矩阵填充问题属于 NP 完全问题^[18]。下节介绍该问题的求解方法。

2.3 MC 问题的常见求解算法

Harm 证明了 MC 问题与 MR 问题实际上是等价的^[22]。这也就意味着最小秩问题的一些求解方法也可用于矩阵填充问题的求解^[11]。假如随机选取秩

为 r 的 $\eta \times n$ 维矩阵 $M \in \mathbb{F}_q$, 参数 ω 为常量 ($2 \leq \omega < 3$)。此外, 定义缺失矩阵元素个数参数 m , 满足

$$m \leq \eta n + r^2 - (\eta + n)r + 1$$

那么目前低秩 MC 问题常见的求解方法主要有以下几点。

1) 暴力破解。对缺失的元素进行穷举赋值, 其时间复杂度为 $q^m r^\omega$ 。

2) 子矩阵求解方法。Coppersmith 等^[25]首先提出这种求解方法, 文献[26]进一步对该类求解方法进行了论证, 但只适用于 $r \ll n$ 的情形。实际意义不大。

3) Shamir 和 Kipnis^[27]提出了 $r \ll n$ 情形的另一种不同求解算法。其思想是将矩阵填充问题转化为一个多变量二次方程组。文献[28]对这种方法做了进一步改进, 其时间复杂度为 $n^{O(r)}$ 。

4) 2000 年, Goubin 等^[28]给出了一种称之为内核攻击的求解算法。当矩阵为方阵 ($n = \eta$) 时, 该算法的时间复杂度约为 $q^{\frac{m}{r}} m^\omega$ 。文献[29]做了进一步改进, 给出了通用求解算法, 其算法的时间复杂度为

$$\min \left(q^{\frac{m}{r}}, q^{\frac{m}{r+m \bmod n}} \right) m^\omega$$

5) 内核攻击方法是 MC 问题的最有效的求解方法。2008 年, Faugere 等^[30]对 Kipnis 和 Shamir 的方法进行了优化改进, 其时间复杂度为 $O(\ln q n^{3(n-r)^2})$ 。需要说明的是, 本文构造的零知识身份认证协议, 其安全参数规模 (包括有限域、矩阵维数及秩等) 的选取依据将主要按照这种内核攻击求解算法。

3 基于 MC 问题的五轮身份认证协议

本节给出五轮身份认证协议的具体构造。同之前的三轮方案^[19]相比, 增加了一次验证者和证明者之间的通信, 增加的挑战值 $k \in (0, q)$, 因此, 单轮攻击者欺骗概率降低至 $\frac{q}{2(q-1)}$, 显然, 当 q 足够大

时, 欺骗概率约等于 $1/2$ 。

3.1 协议构造

本文将该方案分为 2 个阶段。首先是密钥生成阶段, 其次是交互认证阶段。下面分别对这 2 个阶段进行详述。

3.1.1 密钥生成

选定系统参数 η, n, r, m 和 \mathbb{F}_q 。首先, 随机选择一个 $\eta \times n$ 维矩阵 $A = (a_{ij}) \in \mathbb{F}_q$, 且矩阵 A 满足秩 $\text{rank}(A) = r$; 然后, 随机从矩阵 A 中去掉 m 个元素, 将缺失 m 个元素的不完整矩阵记作 A^- , 并用 $(i_1, j_1), \dots, (i_m, j_m)$ 分别表示这 m 个元素在矩阵 A 中的元素位置, i_k 代表矩阵行, j_k 代表矩阵列。

综上所述, 该系统的公钥包含: $\eta, n, r, m, \mathbb{F}_q$ 和 A^- , 私钥为有序序列 s , 满足 $s = \{s_1, \dots, s_m\}$, 其中, $s_k = a_{i_k, j_k}, 1 \leq k \leq m$ 。密钥生成过程即输入满足 $\text{rank}(A) = r$ 的矩阵 A , 进行上述操作后, 输出公私钥对 $(pk, sk) = ((\eta, n, r, m, A^-), s)$ 。

3.1.2 单轮交互方案

证明者随机选择 2 个可逆矩阵 P 和 Q , 其中矩阵 P 和 Q 分别为 $\eta \times \eta$ 、 $n \times n$ 维的方阵, 并随机选取向量 $\gamma \in \mathbb{F}_q^m$ 。证明者机将公钥不完整矩阵 A^- 随机拆分为 2 个元素不完整矩阵之和, 即满足等式 $A^- = A_1^- + A_2^-$, “+”代表相同位置元素相加, 这里与三轮身份认证协议不同的是, 本文指定元素有缺失的矩阵 A_1^- 所有值位置均等于零, 方便后续进行验证。这里需要说明的是这种拆法是公开的, 据此拆分方法验证者同样可以得到残缺矩阵 A_1^- 和 A_2^- 。

证明者接下来秘密地将私钥 s 随机拆成 α 和 β , 满足分量 $\beta_k = \alpha_k + s_k, 1 \leq k \leq m$; 随后将 α 的 m 个元素逐次填充到 A_1^- 缺失元素处得到矩阵 A_1 , 同理, 将 γ 的 m 个元素也填充到 A_2^- 得到矩阵 A_2 。为了后续计算验证以及参数传递, 证明者在本地分别计算 M_1 和 M_2 。其中, 完整矩阵 M_1 将由 s 的 m 个元素依次填充到 A_1^- 得到, 完整矩阵 M_2 将由元素 0 依次填充到 A_2^- 得到, 之后计算矩阵 M , 矩阵 $M = PM_1Q + PM_2Q$, 显然矩阵 M_1 和 M_2 满足 $M_1 + M_2 = A$ 。

五轮协议交互过程如图 1 所示。

证明者和验证者共进行 5 次交互, 整个交互过程及每轮交互中的传递参数的详细描述如下。

1) 证明者先计算矩阵 PA_1Q 和 PA_2Q , 随后计算相应 Hash 值, 分别记作 c_1 和 c_2 , 对应关系如下。

$$c_1 = H(\alpha, M, PA_1Q, PA_2Q)$$

$$c_2 = H(\beta, P, Q, A_2)$$

然后证明者将 $\{c_1, c_2\}$ 发送给验证者。

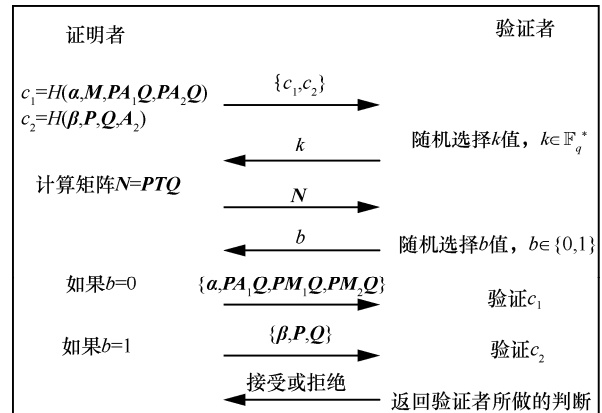


图 1 五轮协议交互过程

2) 验证者从有限域 \mathbb{F}_q 中随机选择一个值 k , 并将其发送给证明者。

3) 证明者计算矩阵 $N = PTQ$ 并发送给验证者, 这里矩阵 T 是通过将 $\gamma + k\alpha + ks$ 的元素逐次填充到 A_2^- 得到的。

4) 验证者随机选择 $b \in \{0, 1\}$ 作为挑战值, 并将 b 传至证明者。

5) 证明者根据 b 值向验证者做出不同的应答。

①如果 $b = 0$, 则证明者将矩阵 PA_1Q 、 PM_1Q 、 PM_2Q 和向量 α 发送至验证者, 验证者收到后计算下述 Hash 值

$$H(\alpha, M, PA_1Q, N - kPM_1Q - kPA_1Q)$$

验证其与 c_1 是否相等, 并计算

$$PM_1Q + PM_2Q = PAQ$$

然后验证相加后得到的矩阵 PAQ 的秩是否等于 r 。

②如果 $b = 1$, 证明者将向量 β 、矩阵 P 和矩阵 Q 发送给验证者, 验证者首先验证 P 和 Q 是否为可逆矩阵, 然后在本地计算下述 Hash 值

$$H(\beta, P, Q, P^{-1}NQ^{-1} - kY)$$

逐次将 β 的元素填充到 A_1^- 的空缺位置得到完整矩阵 Y , 验证 c_2 和该 Hash 值是否相等。

综上所述, 同文献[19]设计的三轮交互协议相比, 多出的一轮交互是验证者选择 k 值并将其传递给证明者; 而证明者收到 k 后, 利用该值计算一个矩阵 N 作为回应。 k 值也是能够将欺骗成功的概率降低到 1/2 的关键。同时在参数生成时, 五轮方案多了一个随机向量 γ , 该向量在验证者验证 Hash 值的时候起到了掩盖私钥的关键作用, 从而确保在

完成验证时并没有泄露私钥。

3.2 安全性证明

下面讨论本文身份认证方案的 3 个重要性质，即完备性、合理性和零知识性。

3.2.1 完备性

定理 1 对于诚实的证明者而言，验证者可以检验证明者的身份。

证明 在证明者和验证者都是诚实的以及不违背交互协议的前提下，证明者因为知道私钥 s 和在第二轮交互中验证者所选的 k 值，很明显可以在每一轮询问中计算出正确的 c_1 和 c_2 值，并返回给验证者相应的信息，进而证明其身份。即针对任何挑战值 b ，证明者做出的应答总能满足以下条件。

1) 当 $b=0$ 时，有如下等式成立

$$H(\alpha, M, PA_1Q, N - kPM_1Q - kPA_1Q) = c_1$$

其中， $PAQ = PM_1Q + PM_2Q$ ；

2) 当 $b=1$ 时，有如下等式成立

$$H(\beta, P, Q, P^{-1}NQ^{-1} - kY) = c_2$$

3) P 、 Q 均为可逆矩阵。

上述矩阵 $N = PTQ$ 中的 T 可由 $\gamma + k\alpha + ks$ 的元素逐次填充至 A_2 得到，矩阵 A_1 和 Y 则分别为逐次将 α 和 β 的元素填充到 A_1 矩阵元素空缺处后得到的完整矩阵。对任意轮次的询问，验证者都能正确检验证明者的身份。综上所述，验证者检验证明者的概率 $\varepsilon = 1$ ，因此，该协议满足完备性。证毕。

3.2.2 合理性

重新设计后的五轮身份认证协议单轮欺骗成功的概率为 $\frac{q}{2(q-1)}$ ，其中 q 为有限域的特征数，

当 q 达到一定大小时，可以将该概率约简为 $1/2$ 。同样，需要重复多次才能达到期望的安全性，以 10^{-6} 安全性为例，文献[19]的三轮交互身份认证协议的欺骗成功概率为 $2/3$ ，达到该安全性需要重复 35 轮，而本文的五轮身份认证协议只需重复 20 轮，便可达到同样的安全性。下面首先对该欺骗成功概率进行分析。

为了能够成功欺骗验证者，恶意证明者 P' 需要做以下准备，以应对验证者可能发起的挑战。同三轮方案不同的是，在传递过程中多了一个 k 值，如果 P' 猜到了正确的 k 值，那么便可以用 k 值来构造相

应的 c_1 和 c_2 ；如果 P' 未猜中对应的 k 值，那么 P' 仍然可以猜测 b 的值来尝试通过验证。将 P' 所做的准备记为 $\{t_1, t_2\}$ 。在 t_1 中， P' 猜测收到的挑战值为 $b'=0$ ， P' 随机生成秩为 r 的矩阵 A' ，故 $\text{rank}(PA'Q) = \text{rank}(A')$ ，即矩阵 $PA'Q$ 的秩为 r ，从而正确计算 c_1 ，而 c_2 可以任意选取。综上所述， P' 可以正确回答挑战 $b=0$ ；在准备 t_2 中， P' 猜测收到的挑战值为 $b'=1$ ， P' 随机生成可逆矩阵 P 和 Q ，以及随机选择向量 α^* ，而 c_1 可以任意选取。由于恶意证明者并不拥有私钥，因此其无法对 2 个挑战值均做出正确回应。因此，在每轮交互过程中 $\{t_1, t_2\}$ 成功的概率为

$$\Pr(\text{Success}) = \Pr(k' = k) +$$

$$\Pr(k' \neq k | b = b = 0) + \Pr(k' \neq k | b = b = 1) =$$

$$\frac{1}{q-1} + \frac{q-2}{q-1} \times \frac{1}{4} + \frac{q-2}{q-1} \times \frac{1}{4} = \frac{q}{2(q-1)}$$

这里假设验证者选取的 k 值不能为 0，否则后续的验证便没有任何意义，所以上述等式中 $k' = k$ 的概率为 $1/(q-1)$ 。不过即使 k 值可以取 0，当 q 达到一定阈值时， $k = k'$ 的概率仍然可以达到约 $1/2$ ，其结果是相同的。通常需要将该身份认证协议重复

w 次，那么全部欺骗成功的概率为 $\left(\frac{q}{2(q-1)}\right)^w$ 。如

果恶意证明者 P' 可以响应验证者所有 b 值的挑战，则有如下定理。

定理 2 假如欺骗者能假冒证明者响应所有 k 值而且能被诚实验证者相信，则存在多项式时间概率图灵机能以不可忽略的概率，要么可以求解矩阵填充问题以恢复私钥，要么可以找到给定哈希函数的一个碰撞。

证明 假如欺骗者能成功欺骗验证者，表明交互过程中欺骗者对所有的挑战均可正确给出响应，可能的挑战有 4 种组合，分别记为 $(k_1, 0), (k_1, 1), (k_2, 0), (k_2, 1)$ ，其中， k_1 代表正确的 k 值。假设 $(P_1A_1Q, P_1M_1Q, P_1M_2Q, \alpha_1, N_1)$ 是对 $(k_1, 0)$ 的应答； (β_1, P_1, Q_1, N_1) 是对 $(k_1, 1)$ 的应答； $(P_2A_1Q_2, P_2M_1'Q_2, P_2M_2'Q_2, \alpha_2, N_2)$ 是对 $(k_2, 0)$ 的应答； (β_2, P_2, Q_2, N_2) 是对 $(k_2, 1)$ 的应答。

上述值均能够通过验证者验证，故 P_1, Q_1, P_2, Q_2 均为可逆矩阵，以及 $P_2M_1'Q_2 + P_2M_2'Q_2$ 和 $P_1M_1Q + P_1M_2Q$ 得到的矩阵的秩均为 r 。且分别有

以下等式成立

$$\begin{aligned} \alpha_1 + A_1^- &= A_1 = \alpha_2 + A_1^- \\ c_1 &= H(\alpha_1, M, P_1 A_1 Q_1, N_1 - k_1 P_1 M_1 Q_1 - k_1 P_1 A_1 Q_1) = \\ &= H(\alpha_2, M', P_2 A_1 Q_2, N_2 - k_2 P_2 M_1' Q_2 - k_2 P_2 A_1 Q_2) \\ M &= P_1 M_1 Q_1 + P_1 M_2 Q_1 \\ M' &= P_2 M_1' Q_2 + P_2 M_2' Q_2 \\ c_2 &= H(\beta_1, P_1, Q_1, P_1^{-1} N_1 Q_1^{-1} - k Y_1) = \\ &= H(\beta_2, P_2, Q_2, P_2^{-1} N_2 Q_2^{-1} - k Y_2) \\ Y_1 &= \beta_1 + A_1^- \\ Y_2 &= \beta_2 + A_1^- \end{aligned}$$

综上所述，证明者要么找到了一个给定 Hash 函数的碰撞，要么就有

$$\begin{aligned} \alpha_1 + A_1^- &= \alpha_2 + A_1^- \\ \beta_1 + A_1^- &= \beta_2 + A_1^- \\ M &= M' \\ N_1 - k_1 P_1 M_1 Q_1 - k_1 P_1 A_1 Q_1 &= \\ N_2 - k_2 P_2 M_1' Q_2 - k_2 P_2 A_1 Q_2 &= \\ P_1^{-1} N_1 Q_1^{-1} - k Y_1 &= P_2^{-1} N_2 Q_2^{-1} - k Y_2 \end{aligned}$$

即找到了矩阵填充问题的一个解。综上所述，该协议满足合理性。证毕。

3.2.3 零知识性

从图 1 可以看出，在验证者和证明者的交互过程中，首先不会造成私钥 s 的泄露。当挑战 $b=0$ 时，向量 α 是随机的， PM_1Q 和 PM_2Q 均为随机矩阵，其相加后得到的 PMQ 为秩为 r 的随机矩阵，不能获得与矩阵 M 有关的任何信息，也就不能获得与私钥 s 有关的任何信息；当挑战 $b=1$ 时，传输矩阵 P 和 Q 是随机选取的，向量 β 也是随机的，故并不会造成私钥 s 泄露。下面给出下述定理的证明，从而指出协议满足零知识性。

定理 3 在随机预言机模型下该身份认证协议(图 1)为零知识交互认证协议。

如果想要证明这一点，需要构造模拟器 U ，满足其输出结果与协议诚实执行后输出是计算不可区分的。下面设计一个模拟器 U 来获取与真实文本计算不可区分的模拟文本，从而证明本文协议的零知识性。

证明 假如模拟器 U 可以跟验证者进行通信。

- 1) U 随机选取一个 b' ， $b' \in \{0,1\}$ 。
- 2) U 随机选取可逆矩阵 P 、 Q 以及向量 α 和 γ ，

U 逐次将 α 和 γ 的元素填充到 A_i^- 得到完整矩阵 A_i^- ，这里 $i \in \{1,2\}$ ，且分别计算 PA_iQ 。若 $b'=0$ ，随机选择秩为 r 的矩阵 A' ，将其拆分为 $M_1 + M_2$ ， $M = PM_1Q + PM_2Q$ ，计算 $c_1 = H(\alpha, M, PA_iQ, PA_2Q)$ ，任取 c_2 即可；如果 $b'=1$ ，则随机选择向量 σ 来代替私钥 s ，并计算 $c_2 = H(\alpha + \sigma, P, Q, A_2)$ ，任取 c_1 即可。

3) U 随机选择一个非零值 $k \in \mathbb{F}_q$ ，并计算矩阵 PTQ 。若 $b=0$ ，矩阵 T 可通过计算将 $PA_2Q + kPM_1Q$ 得到，这里 A_2 由将 $\gamma + k\alpha$ 的元素逐次填充到 A_2^- 得到；如果 $b=1$ ，矩阵 T 可通过将 $\gamma + k\alpha + k\sigma$ 的元素逐次填充到 A_2^- 得到。

4) 模拟器 U 随机选择一个 $b \in \{0,1\}$ ，若 $b \neq b'$ ，则直接返回步骤 1) 重新执行；若 $b = b'$ ，则记录该轮交互过程。若 $b = b' = 0$ ，则 U 返回的信息 $\{c_1, c_2, k, N, \alpha, A', PA_iQ, PM_1Q, PM_2Q\}$ 是与真实值是计算不可区分的；若 $b = b' = 1$ ，则模拟器 U 返回的 $\{c_1, c_2, k, N, P, Q, \alpha + \sigma\}$ 是与真实文本计算不可区分的。由于模拟器 U 可以同验证者进行任意次数的通信，因此模拟器可以重启整个认证协议直到猜对 b 值为止。

综上所述，该协议满足零知识性。

4 性能分析及方案比较

对于本文设计的五轮零知识身份认证方案，其安全性基于矩阵填充问题，所以根据 2.3 节中的描述可知，当选取 11×11 维的矩阵时，本文五轮身份认证协议具有 80 bit 以上安全性，可以满足目前实际应用需求。因此本文身份认证方案在实际应用中推荐矩阵参数为有限域 \mathbb{F}_{65521} 上的 11 维方阵。

交互过程的前四轮较为简单，可以看出传递的参数为 2 个 Hash 值以及 k 值和 b 值（其中 $k \in \mathbb{F}_q$ ， $b \in \{0,1\}$ ），以及矩阵 N ，所以只需要传递 $(2 \times 160 + lbq + \eta n lbq + 1)$ bit，其中 q 为有限域的特征数。同样，这里为了方便同其他方案做比较，故用于计算的 Hash 值的长度为 160 bit。

由上节中协议交互过程易知，当 $b=0$ 时，传递的数据为向量 α 以及其他 3 个维数为 $\eta \times n$ 的矩阵，向量 α 有 m 个元素，共计为 $(3\eta n lbq + m lbq)$ bit；当 $b=1$ 时，需传递矩阵 P 、 Q 以及向量 β ，共计 $(\eta^2 lbq + n^2 lbq + m lbq)$ bit，这里 m 为矩阵中缺失元素

表 1 参数比较

参数	PKP ^[13]	PPP ^[16]	SC ^[17]	MC(三轮) ^[19]	MC(五轮)
矩阵维数	16 × 34	101 × 117	64 × 2 048	11 × 11	11 × 11
有限域	\mathbb{F}_{251}	\mathbb{F}_2	\mathbb{F}_{257}	\mathbb{F}_{65521}	\mathbb{F}_{65521}
通信次数	5	5	5	3	5
轮次	20	35	20	35	20
欺骗成功概率	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{1}{2}$
系统参数/kbit	4.6	28.5	1 048	0.7	0.7
公钥尺寸/bit	272	11 918	512	432	432
私钥尺寸/bit	128	117	2 048	144	144
通信开销/bit	665	1 040	33 216	4 449	7 257
时间复杂度	2 ⁶⁰	2 ⁷⁴	2 ¹⁰⁰	2 ⁸⁸	2 ⁸⁸

的个数。综上所述，整个交互过程所需要的比特数平均为

$$2 \times 160 + 1 + \frac{5}{2} \eta n l b q + \frac{1}{2} (\eta^2 + n^2) l b q + (m + 1) l b q$$

3.2.2 节中分析过，在 q 达到一定阈值时，单轮交互过程中的欺骗成功概率为 $1/2$ ，所以如果想要达到 10^{-6} 的安全性实际需求，通常至少需要交互 20 轮以上。

表 1 列举了 80 bit 安全性水平下，本文方案与其他方案的比较结果，其中 MC（五轮）代表本文所提方案。很显然，本文方案和 MC（三轮）方案的公钥尺寸较小。通过比较可得，本文方案在公私钥以及参数尺寸并未发生变化的前提下，通过增加一轮交互，降低恶意证明者成功通过验证者身份认证的概率至 $1/2$ 。同时单轮方案的通信开销虽然由于方案设计改变验证参数导致有所增加，但由于欺骗成功概率的降低，使达到相同安全性时交互轮次降低，故本文设计的五轮交互身份认证方案与 MC（三轮）方案相比，总的通信开销相差不大，但进一步提高了协议的安全性。

本文对几种典型的身份认证方案进行了实验比较，实验结果如表 2 所示，本文方案实现效率比 Stern 方案^[7]和文献[19]方案要高，仅需 4.6 ms。同时也不难发现，由于本文身份认证协议需要执行多轮，相较于基于数论困难问题设计的 1 024 bit GQ 方案^[3]和 512 bit 的 Schnorr 方案^[4]，其实现效率略低一些，但在实际工程应用中仍然在可接受的范围，此外本文方案还具有抗量子计算攻击的优势。

表 2 实现效率比较

方案	问题	时间/ms
GQ 方案 ^[3]	IFP	2
Schnorr 方案 ^[4]	DLP	3.1
Stern 方案 ^[7]	SDP	29.3
文献[19]方案	MCP	5.8
本文方案	MCP	4.6

注：实验环境为 Intel i5-5200U 2.2 GHZ/4 GB。

5 结束语

本文基于（低秩）矩阵填充问题提出了一种五轮身份认证方案，相比已有的三轮方案，本文所提方案将单轮交互方案的欺骗成功概率从 $2/3$ 降低到 $1/2$ ，并给出了欺骗概率的详细分析及严格安全性证明。以 10^{-6} 安全性为例，五轮方案只需要重复执行 20 次，而三轮方案需要执行 35 次。所提身份认证方案具有很好的抗量子计算攻击潜力。而且基于本文提出身份认证方案，通过 Fiat-Shamir 密码转换技术，还可得到具有抗量子计算性质的数字签名方案^[31-32]。

另外，能否在本文身份认证协议的基础上，优化参数降低通信开销，从而进一步提高协议的执行效率，值得进一步研究。

参考文献：

[1] GOLDWASSER S, MICALI S, RACKOFF C. The knowledge complexity of interactive proof systems[J]. SIAM Journal on Computing, 1989, 18(1): 186-208.
 [2] FEIGE U, FIAT A, SHAMIR A. Zero-knowledge proofs of identity[J].

- Journal of Cryptology, 1988, 1(2): 77-94.
- [3] GUILLOU L C, QUISQUATER J J. A “paradoxical” identity-based signature scheme resulting from zero-knowledge[C]//Advances in Cryptology — CRYPTO’88. Berlin: Springer, 1990: 216-231.
- [4] SCHNORR C P. Efficient signature generation by smart cards[J]. Journal of Cryptology, 1991, 4(3): 161-174.
- [5] SHOR P W. Algorithms for quantum computation: discrete logarithms and factoring[C]//Proceedings of the 35th Annual Symposium on Foundations of Computer Science. Piscataway: IEEE Press, 1994: 124-134.
- [6] ALAGIC G, ALPERIN-SHERIFF J, APON D, et al. Status report on the first round of the nist post-quantum cryptography standardization process[R]. 2019.
- [7] STERN J. A new identification scheme based on syndrome decoding[C]//Advances in Cryptology — CRYPTO’93. Berlin: Springer, 1994: 13-21.
- [8] GABORIT P, GIRAULT M. Lightweight code-based identification and signature[C]//Proceedings of 2007 IEEE International Symposium on Information Theory. Piscataway: IEEE Press, 2007: 191-195.
- [9] AGUILAR C, GABORIT P, SCHREK J. A new zero-knowledge code based identification scheme with reduced communication[C]//Proceedings of 2011 IEEE Information Theory Workshop. Piscataway: IEEE Press, 2011: 648-652.
- [10] CAYREL P L, EL Y A S M, HOFFMANN G, et al. An improved threshold ring signature scheme based on error correcting codes[C]//Arithmetic of Finite Fields. Berlin: Springer, 2012: 45-63.
- [11] COURTOIS N T. Efficient zero-knowledge authentication based on a linear algebra problem MinRank[C]//Advances in Cryptology — ASIACRYPT 2001. Berlin: Springer, 2001: 402-421.
- [12] SAKUMOTO K, SHIRAI T, HIWATARI H. Public-key identification schemes based on multivariate quadratic polynomials[C]//Advances in Cryptology — CRYPTO 2011. Berlin: Springer, 2011: 706-723.
- [13] SHAMIR A. An efficient identification scheme based on permuted kernels (extended abstract)[C]//Advances in Cryptology — CRYPTO’89 Proceedings. Berlin: Springer, 1989: 606-609.
- [14] CHEN K F. A new identification algorithm[C]//Cryptography: Policy and Algorithms. Berlin: Springer, 1996: 244-249.
- [15] STERN J. Designing identification schemes with keys of short size[C]//Advances in Cryptology — CRYPTO’94. Berlin: Springer, 1994: 164-173.
- [16] POINTCHEVAL D. A new identification scheme based on the perceptrons problem[C]//Advances in Cryptology — EUROCRYPT’95. Berlin: Springer, 1995: 319-328.
- [17] CAYREL P L, LINDNER R, RÜCKERT M, et al. Improved zero-knowledge identification with lattices[C]//Provable Security. Berlin: Springer, 2010: 1-17.
- [18] YANG R P, AU M H, ZHANG Z F, et al. Efficient lattice-based zero-knowledge arguments with standard soundness: construction and applications[C]//Advances in Cryptology — CRYPTO 2019. Cham: Springer International Publishing, 2019: 147-175.
- [19] 王后珍, 郭岩, 张焕国. 基于矩阵填充问题的高效零知识身份认证方案[J]. 武汉大学学报(理学版), 2021, 67(2): 111-117.
WANG H Z, GUO Y, ZHANG H G. Efficient zero-knowledge identification based on matrix completion problem[J]. Journal of Wuhan University (Natural Science Edition), 2021, 67(2): 111-117.
- [20] BUSS J F, FRANSEN G S, SHALLIT J O. The computational complexity of some problems of linear algebra[J]. Journal of Computer and System Sciences, 1999, 58(3): 572-596.
- [21] GABIDULIN E M. Theory of codes with maximum rank distance (translation)[J]. Problems of Information Transmission, 1985, 21(1): 1-12.
- [22] HARM D. On the equivalence between low-rank matrix completion and tensor rank[J]. Linear and Multilinear Algebra, 2018, 66(4): 645-667.
- [23] PEETERS R. Orthogonal representations over finite fields and the chromatic number of graphs[J]. Combinatorica, 1996, 16(3): 417-431.
- [24] CRAVO G. Matrix completion problems[J]. Linear Algebra and Its Applications, 2009, 430(8/9): 2511-2540.
- [25] COPPERSMITH D, STERN J, VAUDENAY S. Attacks on the birational permutation signature schemes[C]//Advances in Cryptology — CRYPTO’93. Berlin: Springer, 1993: 435-443.
- [26] COURTOIS N T. The security of hidden field equations (HFE)[C]//Topics in Cryptology — CT-RSA 2001. Berlin: Springer, 2001: 266-281.
- [27] KIPNIS A, SHAMIR A. Cryptanalysis of the HFE public key cryptosystem by relinearization[C]//Advances in Cryptology — CRYPTO’99. Berlin: Springer, 1999: 19-30.
- [28] GOUBIN L, COURTOIS N T. Cryptanalysis of the TTM cryptosystem[C]//Advances in Cryptology — ASIACRYPT 2000. Berlin: Springer, 2000: 44-57.
- [29] COURTOIS N T. The security of cryptographic primitives based on multivariate algebraic problems: MQ, MinRank, IP, HFE[D]. Paris: Paris 6 University. 2001.
- [30] FAUGERE J C, LEVY-DIT-VEHEL F, PERRETL. Cryptanalysis of minrank[C]//Advances in Cryptology-CRYPTO’08. Berlin: Springer, 2008: 280-296.
- [31] 张海波, 黄宏武, 刘开健, 等. 车联网中可证安全的匿名可追溯快速组认证协议[J]. 通信学报, 2021, 42(6): 213-225.
ZHANG H B, HUANG H W, LIU K J, et al. Verifiably secure fast group authentication protocol with anonymous traceability for Internet of vehicles[J]. Journal on Communications, 2021, 42(6): 213-225.
- [32] 田苗苗, 陈静, 仲红. 格上基于身份的增量签名方案[J]. 通信学报, 2021, 42(1): 108-117.
TIAN M M, CHEN J, ZHONG H. Identity-based incremental signature scheme from lattices[J]. Journal on Communications, 2021, 42(1): 108-117.

[作者简介]



王后珍(1981-), 男, 湖北恩施人, 博士, 武汉大学讲师, 主要研究方向为信息安全、抗量子密码、量子计算等。

蔡鑫伟(1998-), 男, 湖北武汉人, 武汉大学硕士生, 主要研究方向为信息安全、应用密码学等。

郭岩(1998-), 男, 河北邢台人, 武汉大学硕士生, 主要研究方向为信息安全、应用密码学等。

张焕国(1945-), 男, 湖北武汉人, 博士, 武汉大学教授, 主要研究方向为信息安全、密码学、可信计算等。